

A B S T R A C T

A temporary model generating unit (103) generates a probability model which is estimated to generate a text document as a processing target and in which information indicating which word of the text document to which topic is made to correspond to a latent variable, and each word is made to correspond to an observable variable. A model parameter estimating unit (105) estimates model parameters defining a probability model on the basis of the text document as the processing target. When a plurality of probability models are generated, a model selecting unit (107) selects an optimal probability model on the basis of the estimation result for each probability model. A text segmentation result output unit (108) segments the text document as the processing target for each topic on the basis of the estimation result on the optimal probability model.

This saves the labor of adjusting parameters in accordance with the characteristics of a text document as a processing target, and eliminates the necessity to prepare a large-scale text corpus in advance by spending much time and cost. In addition, this makes it possible to accurately segment a text document as a processing target independently of the contents of the document, i.e., the domains.